基于时间序列与 BP-ANN 的短时交通流速度预测模型研究 *

田瑞杰，张维石，翟华伟†

(大连海事大学 信息科学技术学院，辽宁 大连 116026)

摘　要： 短时交通流速度预测被认为是智能交通系统的重要组成部分，预测模型的准确性在一定程度上决定了实时交通控制和管理的性能。针对现有的交通流速度预测模型使用唯一数据集且模型单一的问题，提出一种时间序列与人工神经网络相结合的预测模型。该模型通过时间序列分别对实时数据和历史数据建模预测，并应用人工神经网络调整实时数据和历史数据的预测值。实验结果表明该预测模型能够将预测误差控制在 7%以内，且能够对不同输入参数下的短时交通流速度进行有效预测。

关键词： 时间序列；人工神经网络；短时预测；交通流速度

中图分类号： TP391.9　　doi: 10.3969/j.issn.1001-3695.2018.06.0308

Short-term traffic flow velocity prediction model based on time series and BP-ANN

Tian Ruijie, Zhang Weishi, Zhai Huawei†

(*Information Science & Technology, Dalian Maritime University, Dalian Liaoning* 116026, *China*)

Abstract: Short-term traffic flow velocity prediction is considered as an important part of the intelligent transportation system. The accuracy of the prediction model determines the performance of real-time traffic control and management to some extent. Aiming at the problem that the existing traffic flow velocity prediction model uses a unique data set and a single model, this paper proposes a prediction model combining time series and artificial neural network. The model predicts real-time data and historical data by time series, and uses artificial neural network to adjust the predicted values of real-time data and historical data. The experimental results show that the prediction model can control the prediction error within 7% and can effectively predict the short-term traffic flow speed under different input parameters.

Key words: time series; artificial neural network; short-term prediction; traffic flow velocity

0　引言

交通预测已成为智能交通系统中的一个重要课题错误!未找到引用源。，其在日常生活中扮演着重要角色。交通道路预测是现在人们广泛使用的智能车辆路线导航设备、电子地图等工具的重要功能之一，特别是短时交通道路预测，通过对道路状况进行实时预测分析，导航或电子地图等工具能够帮助使用者尽可能避开拥堵路段，提供更加合理和多样的出行路线选择。宏观交通运行特性指标为流量、密度和速度错误!未找到引用源。，而在公交车道效益评价方面，最重要的评价指标包括行程车速、通道客运能力、道路使用者出行时间节省等错误!未找到引用源。，因此本文选取交通流速度指标来反映公交车道路状态。通过时间序列模型来分析时序数据的线性相关性，并将预测的结果结合预测范围、预测时间作为人工神经网络的输入来完成模型的分析预测。

1　国内外相关研究

近年来关于短时交通流速度预测产生许多研究成果，研究方法大致分为两类，即参数方法和非参数方法。典型的参数方法是使用时间序列模型，例如 Lai **错误!未找到引用源。**等人利用时间序列模型和线性回归模型进行高速铁路短期客流预测，其主要贡献是将两种模型结合起来，利用不同的模型处理不同的数据部分，通过最小绝对值方法进行整合以生成最终的预测模型，从而提高预测精度，但本文主要侧重于城市交通流速度预测。对于非参数方法，Moreno 错误!未找到引用源。等人采用模糊 C-均值聚类的算法对道路交通流速度进行估计，对接收到的瞬时速度数据进行聚类划分等级，然后得到一个隶属度矩阵与聚类中心向量，最后获取平均速度，该研究并没有考虑公交车进出站时的车辆非正常行驶的瞬时速度变化对聚类结果的影响，这

收稿日期： 2018-06-18；**修回日期：** 2018-07-27　　**基金项目：** 中央高校基本科研基金资助项目（3132016308，3132018197）；辽宁省自然科学基金资助项目（20170520196）

作者简介： 田瑞杰（1992-），男，山西忻州人，硕士研究生，主要研究方向为计算机软件与理论、智能交通、大数据；张维石（1962-），男，教授，博导，主要研究方向为软件工程与软件理论、交通信息系统等；翟华伟（1979-），男（通信作者），讲师，博士，主要研究方向为软件工程与软件方法学、智能交通、大数据应用等（zhw@dlmu.edu.cn）.

会严重影响预测精度。Cao 错误!未找到引用源。设计了基于 BP 神经网络的城市干道交通流预测，该研究使用历史数据作为神经网络训练集和测试集，在估计交通流时，只体现出道路交通流的周期性和普遍性，并没有结合实时道路状况进行分析，从而未体现出道路交通流的特殊性，数据单一是该研究模型的不足之处，其次，Zhang 错误!未找到引用源。指出，神经网络模型在处理大量具有非线性特征的历史数据时具有优势，但忽略了包括数据在内的线性关系。对此，Li 等人错误!未找到引用源。设计了短期交通流预测模型，该模型结合了 ARIMA 和径向基函数人工神经网络(RBF-ANN)，利用 ARIMA 模型对交通流时间序列的线性分量进行建模，然后利用 RBF-ANN 模型通过 ARIMA 模型建模的残差来捕获非线性分量，但缺点是该研究并没有对高峰段和非高峰段的交通流做研究，因此当交通路况处于高峰拥堵时，交通状况的急剧改变会严重影响模型的预测性能。

综上，短时交通流速度预测存在不确定性较强和规律性较弱等一系列不可忽略的特点，单一使用某一种模型进行预测计算，很难进一步提高短时交通流速度预测的准确性。为此，本文提出了一种时间序列(time series)与人工神经网络(ANN)相结合的模型 time series-ANN（简称 TS-ANN），以克服单一模型的局限性，让两者进行优势互补，构建出完善的短时交通流速度预测模型。在 TS-ANN 下预测交通流速度是发展智能城市的一次重要尝试。

2 TS-ANN 模型

在本章提出 *TS-ANN* 预测模型，图 1 显示了模型的结构。实时数据 $\bar{V}_{实时}=\{\bar{v}_1, \bar{v}_2, \bar{v}_3, \ldots, \bar{v}_m\}$表示路段平均速度(*km/h*)实时信息，历史数据 $\bar{V}_{历史}=\{\bar{v}_1, \bar{v}_2, \bar{v}_3, \ldots, \bar{v}_n\}$表示路段平均速度(*km/h*)历史信息，分别从 $\bar{V}_{实时}$，$\bar{V}_{历史}$中选取一个子集 $V_r=\{\bar{v}_i, \bar{v}_{i+1}, \ldots, \bar{v}_j\}(i\geq 1, j\leq m)$，$V_h=\{\bar{v}_i, \bar{v}_{i+1}, \ldots, \bar{v}_j\}(i\geq 1, j\leq n)$作为时序数据进行时间序列建模分析，从而预测出下一时刻路段的平均速度，其中 V_r 中相邻元素时间间隔 $\Delta t=1$ 分钟，V_h 中相邻元素时间间隔 $\Delta t=1$ 周。时间序列建模时首先对时序数据进行平稳性检验，其次观察自相关值 *ACF* 和偏自相关值 *PACF*，从 *AR*(*p*)，*MA*(*q*)，*ARMA*(*p*,*q*)中选择最满足时序数据的模型，进而根据赤池信息准则(*AIC*)，施瓦兹准则(*SC*)，汉南-奎因准则(*HQC*)确定滞后阶数 *p*、*q* 值，最后通过检测残差序列是否为高斯白噪声来对模型进行拟合检验。在 *TS-ANN* 模型中，预测范围和预测时间被认为会影响预测模型的预测精度（2.1.3 节详细讨论），预测范围单位是分钟，预测时间为当前时间与预测范围之和，由于本文的预测时间共有 2 种状态为高峰段和非高峰段，所以预测时间设置为 *time*{1,2}，1 表示高峰段，2 表示非高峰段。时间序列预测值 $V_{实时}$，$V_{历史}$结合预测范围、预测时间作为人工神经网络的输入，*TS-ANN* 模型中人工神经网络选用 *BP* 神经网络。

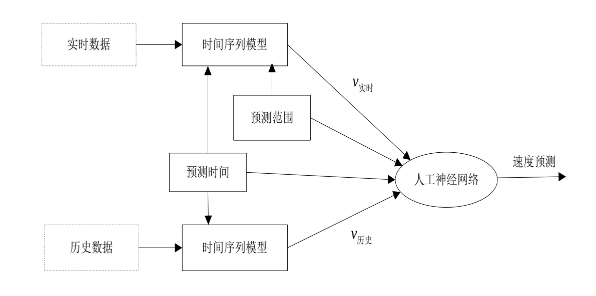


图 1 TS-ANN 预测模型结构

BP 网络通常包含一个输入层，一个或多个隐含层和一个输出层。相邻层的节点彼此连接，并且仅与每个节点连接接收前一层神经元的输入，只有每层神经元处理的输入信息才能成为输出层的输出。

假设输入层，隐含层和输出层分别具有 *m* 个神经元，*h* 个神经元和 *p* 个神经元, 隐含层的输入和输出用等式(1)和(2)表示，输出层的输入和输出用公式(3)和(4)表示。

$$I_j = \sum w_{ji} * x_i + b_j (i = 1, 2, 3, \ldots, m; j = 1, 2, 3, \ldots, h) \quad (1)$$

$$y_j = f_h(I_j)(j = 1, 2, 3, \ldots, h) \quad (2)$$

$$I_o = \sum w_{oj} * y_j + b_o (j = 1, 2, 3, \ldots, h; o = 1, 2, 3, \ldots, p) \quad (3)$$

$$y_o = f_p(I_o)(o = 1, 2, 3, \ldots, p) \quad (4)$$

其中：w_{ji} 表示输入层和隐含层的连接权重，w_{oj} 表示隐含层和输出层的连接权重，x_i 表示输入层的输入值，I_j 和 I_o 分别表示隐含层和输出层的输入值，b_j 和 b_o 是隐含层和输出层的阈值，f_h 和 f_p 分别是隐含层和输出层的激活函数。通常，隐含层中常用的激活函数是 *tanh* 或 *sigmod*，输出层的激活函数是 *relu*，相应的公式如下：

$$f(x) = \frac{2}{1+e^{-2x}} - 1 \quad (5)$$

$$f(x) = \frac{1}{1+e^{-x}} \quad (6)$$

$$f(x) = \max(0, x) \quad (7)$$

在算法 1 中，首先随机初始化权重值 *w* 阈值 *b*，参数的维数由输入层、隐含层、输出层节点个数决定。第一次 *for* 循环时，分别对实时数据和历史数据进行时间序列建模预测，两个模型的预测结果结合预测范围和预测时间被传入到第二个 *for* 循环，结合预测时间和预测范围作为 *ANN* 模型的输入，输入数据为{$V_{实时}$,$V_{历史}$, *time*, *PH*}，训练 *ANN* 模型，比较神经网络的预测结果与现实世界的交通数据，计算损失函数并应用反向传播算法来学习参数，训练模型在算法 1 结束时返回。

在算法 2 中，时间序列模型用于生成第一步预测结果。再结合预测范围和预测时间作为第二步预测的输入，以产生 *TS-ANN* 的最终结果，输出数据格式为 $V_{预测}$。

算法 1 *TS-ANN* 模型下训练 *ANN*

输入：实时数据集 $V = \{V_i(t)\}$

历史数据集 $V = \{V_j(t)\}$

$T=\{t_1,\ldots t_m\}$，*t* 表示每个速度信息的时间值

预测范围集 $PH=\{h_1,h_2,\ldots,h_n\}$

输出：*ANN* 训练模型

开始：初始化连接权重 $w_{hi}\in R^{4*n}$，$w_{oi}\in R^n$, 阈值 $b^h\in R^n$，$b^o\in R^1$

（*n* 为隐含层神经元数量）

for(k=1;$k\le m$;k++)

for(g=1;$g\le n$;g++)

使用 $V=\{V_i(t)\}$和 t_k，d_k，h_g 进行时间序列模型建模产生预测值 V_R；

使用 $V=\{V_j(t)\}$和 h_g 进行时间序列模型建模产生预测值 V_H；

for(k=0;k<$train_num$;k++)

将预测值 V_R，V_H，以及预测时间、预测范围输入到 *ANN* 模型，使用梯度下降来学习参数

训练 *ANN* 模型，

使用损失函数评估模型的性能

返回：*ANN* 训练模型

算法 2 *TS-ANN* 下的路段速度估计

输入：实时数据集 $V=\{V_i(t)\}$

历史数据集 $V=\{V_j(t)\}$

预测时间 *time*

预测范围 *PH*

输出：*TS-ANN* 模型预测结果

开始：初始化连接权重 $w_{hi}\in R^{4*n}$，$w_{oi}\in R^n$, 阈值 $b^h\in R^n$，$b^o\in R^1$ 为算法 1 的输出值（*n* 为隐含层神经元数量）

实时数据集 *V* 进行时间序列建模产生预测值 V_R

历史数据集 *V* 进行时间序列建模产生预测值 V_H

将 V_R，V_H，*time*，*PH* 输入到 *ANN* 模型

计算最终预测结果

返回：预测结果

2.1 时间序列预测模型

基本模型是 *ARMA(p,q)*，表达式为

$$Y_t=\sum_{i=1}^{p}\varphi_i Y_{t-i}+\sum_{j=0}^{q}\theta_j\varepsilon_{t-j} \tag{8}$$

其中：Y_t(t=1,2,3…)为时间序列，ε 是零均值白噪声，φ 为 *AR* 模型参数，θ 为 *MA* 模型参数，*p* 为自回归项，*q* 为移动平均项数，*t* 是时间索引。*ARMA(p,q)*是平稳随机过程**错误!未找到引用源。**。但从现实世界获得的时序数据经常会呈现出系统性地上升或下降等趋势，有些还呈现出周期性波动，这样的时序数据产生于非平稳的随机过程，不可以直接使用 *ARMA(p,q)*建模分析。对于非平稳的时间序列，首先应用差分变换将其平稳化，然后再使用 *ARMA(p,q)*来模拟已平稳化的随机过程，这就是自回归积分滑动平均模型，简记为 *ARIMA(p,d,q)*, 表达式为

$$\varphi_p(B)(1-B)^d Y_t=\theta_q(B)\varepsilon_t \tag{9}$$

其中：*B* 是后移算子，其中 $BY_t=Y_{t-1}$，*d* 是差分次数。

2.1.1 时序数据集构造

本文从大连市交通数据库获取路段 *L9001*，*L9002*，*L9003*，*L9004*（如图 2）上二个月公交车轨迹数据，时间起讫点为 2017 年 11 月 1 日至 2017 年 12 月 29 日（每天上午 6 点至晚上 9 点），时间间隔为 10 s，对应的 4 个路段的路程分别为 *0.84km*、*0.98km*、*0.85km*、*0.97km*，表 1 为部分轨迹数据。

针对采集的轨迹数据，使用路段平均速度估计算法，从而得到本文实验数据集 *V(t)*，其中 $V_{L900i}(t)=\{V_1(t), V_2(t)\}$(i=1,2,3,4)，实时数据 $V_1(t)$是某天某时刻之前的连续 p_1 个时刻的路段平均速度信息，表达式为

$$v_1(t)=\{v(t-1),v(t-2),\ldots,v(t-p_1)\} \tag{10}$$

历史数据 $V_2(t)$是某天某时刻之前的连续 p_2 周该时刻的路段平均速度信息，表达式为

$$v_2(t)=\{v(t-7\times24\times1),v(t-7\times24\times2),\ldots,v(t-7\times24\times p_2)\} \tag{11}$$

表 1 轨迹数据示例

carid	coordinate		speed/km/h	heading	course/km	datetime
	lon	lat				
B001	121.589	38.93	10	134	0.8	2017-11-1 7:25:20
B002	121.573	38.93	15	133	0.24	2017-11-20 7:48:58
B001	121.570	38.98	11	358	0.07	2017-12-10 7:48:58
…	…	…	...	…	…	…
B002	121.598	38.91	32	85	0.67	2017-12-29 7:25:20

表中：*carid* **表示**车辆编号；*lon* **表示**车辆定位经度；*lat* 表示车辆定位纬度；*speed* **表示**车辆瞬时速度；*heading* **表示**车辆行驶方向；*course* 表示与上一站距离；*datetime* 表示数据上传时刻。

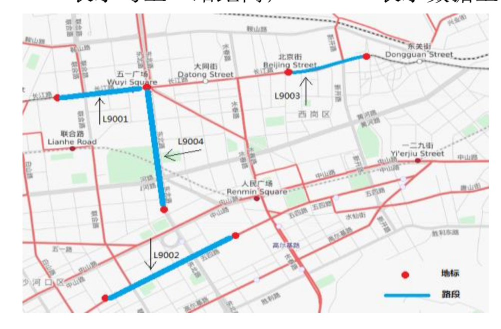


图 2 所选道路位置

2.1.2 时间序列模型构建

本节选取路段 *L9001* 的实时数据 $V_1(t)$和历史数据 $V_2(t)$进行实验分析，部分输入数据如表 2、3 所示。

表 2 实时数据示例

时间	速度(*km/h*)
2017-12-29 6:30	43.6
2017-12-29 6:31	42.4
2017-12-29 6:32	45.7
...	...
2017-12-29 9:59	31.3

2017-12-29 10:00	28.7

表 3　历史数据示例

时间	速度(*km/h*)
2017-12-29 8:00	30.4
2017-12-22 8:00	26.5
2017-12-15 8:00	27.1
...	...
2017-12-8 10:00	31.2
2017-12-3 10:00	28.6

1）单位根检验

对路段 *L9001* 的时序数据 $V_1(t)$进行平稳性检验[错误!未找到引用源。]，如表 4 所示，当检验式中的附加项为 *Intercept* 时，在原假设 $H_0:\delta=0$ 下，单位根的 *t* 统计量的值为-3.6492，在 1%，5%，10%三个显著性水平下，单位根检验的临界值分别为-3.5683，-2.9211，-2.5985，显然，ζ 检验统计量值小于相应的 *DW* 临界值，从而拒绝原假设 H_0，表明该序列是平稳序列，可以用于建模分析。

表 4　单位根检验信息表

Include in test equation	t-Static			
	ADF test statistic	Test critical values		
		1% level	5% level	10% level
None	-1.3864	-2.6140	-1.9478	-1.6124
Intercept	-3.6492	-3.5683	-2.9211	-2.5985
Trend and Intercept	-3.7196	-4.1525	-3.5023	-3.1806

2）模型选择

观察时序数据 $V_1(t)$的 *ACF/PACF*，*ACF* 是具有特定时间偏移的时间序列中的任何两个值之间的相关性，称为滞后(*lag*)，*PACF* 是具有特定滞后的任意两点之间的相关性[错误!未找到引用源。]，结合自回归模型和移动平均模型的定义，可以确定时间序列模型的 *ACF/PACF* 表现，如表 5 所示。通过观察 *ACF/PACF* 能够选择模型的参数数量，如图 3 所示，*ACF* 表现出 *lag=2* 时拖尾，*PACF* 表现出 *lag=2* 时拖尾，所以初步确定该模型是 *ARMA* 模型，并且将 *p* 的值限定在区间[1,2]，*q* 的值限定在区间[1,2]。

表 5　时间序列模型 *ACF/PACF* 表现

	AR(p)	*MA(q)*	*ARMA(p,q)*
ACF	拖尾	*q* 阶截尾	拖尾
PACF	*p* 阶截尾	拖尾	拖尾

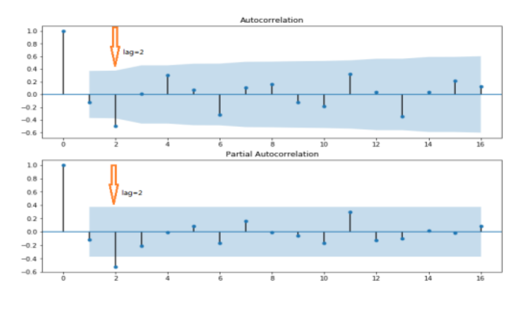


图 3　自相关和偏自相关分析

3）滞后阶数选择

使用 *AIC, SC, HQC* 信息准则检验 *ARMA* 模型拟合度[错误!未找到引用源。]，从而确定滞后阶数 *p, q* 的值。根据 *p, q* 的取值区间建立 *ARMA*(2,2)、*ARMA*(2,1)、*ARMA*(1,2)、*ARMA*(1,1)四种模型，对于每个模型，一共进行三次模拟实验，每次实验的样本数量分别设置为 40, 50, 60, 选择三次实验中 *AIC*、*SC*、*HQC* 的最小值。从表 6 知，选择 *ARMA*(1,2)作为最终模型，因为它具有最小的 *AIC*、*SC*、*HQC* 值，模型预测效果最好。

表 6　不同 *ARMA* 模型的 *AIC SC HQC*

Model	*AIC*	*SC*	*HQC*
ARMA(2,2)	6.895573	7.088580	6.968777
ARMA(2,1)	7.020063	7.174497	7.078665
ARMA(1,2)	6.888170	7.041132	6.946418
ARMA(1,1)	7.037046	7.151767	7.080732

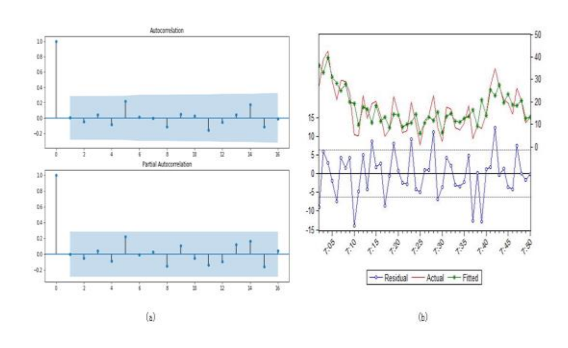


图 4　(*a*)残差序列 *ACF* 和 *PACF*　(*b*)*ARMA*(1,2)预测值和实际值拟合

4)模型检验

评估模型是否能很好地表示时间序列，应用了模型检验方法。该模型对残差序列进行相关性和偏自相关性检验，如果模型与时间序列很好的拟合，则残差部分表现为白噪声序列。如图 4(*a*)，计算残差序列的 *ACF* 和 *PACF* 值，根据滞后的阶数确定残差序列已为白噪声序列。图 4(*b*)显示了模型预测值与实际值拟合效果，可以看出，在进入早高峰期时，拟合效果明显降低，猜测是复杂的交通状况引起交通流速度的急剧变化，从而使模型的拟合性能下降。重复以上步骤，时序数据 $V_2(t)$采用 *ARIMA*(2,1,2)模型。

2.1.3 模型分析

该节通过设置不同预测范围(*PH*)和预测时间(*time*)对模型的性能进行分析研究。

假设 1　当预测范围不超过一定阈值 φ，即 $PH<\varphi$，*ARMA*(1,2)模型预测效果更为准确，当预测值超过一定阈值 φ，即 $PH>\varphi$，用 *ARIMA*(2,1,2)模型预测效果更为准确。

如图 5 所示，使用平均绝对百分比误差 *MAPE*（公式 12）来评价预测的准确性，当 $PH<5$ 时（即预测范围小于 5 分钟），*ARMA*(1,2)模型有更好的预测效果。然而，当 $PH>5$（预测范围增加到大于 5 分钟时），*ARIMA*(2,1,2)模型开始产生更好的预测效果。设置 $\varphi=5$，假设 1 成立。

$$MAPE = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{|X_i - Y_i|}{X_i}\right)\times 100\% \quad (12)$$

其中：X_i 和 Y_i 分别代表实际交通速度和预测交通速度，*n* 表示预测次数。

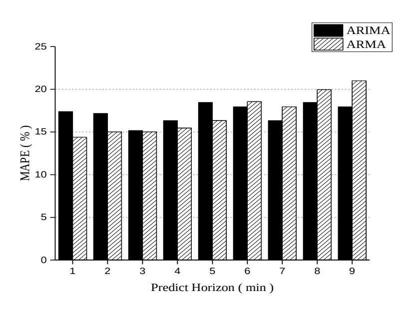


图 5　不同 *PH* 下模型 *MAPE* 值对比

假设 2　*ARIMA*(2,1,2)模型可以高效地预测高峰时段的交通流速度变化，而 *ARMA*(1,2)模型很难捕捉道路状态急剧变化特征。

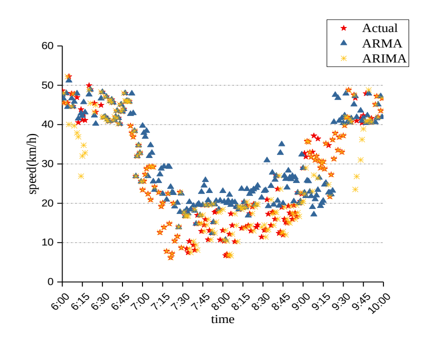


图 6　实际值与时间序列模型输出结果比较

在该案例中，固定预测范围 *PH=5*(*φ=5*)，图 6 描述了特定工作日不同时间的实际速度值和 *ARMA*(1,2)、*ARIMA*(2,1,2)的预测值。在凌晨 6:50 左右的早晨高峰期，*ARIMA*(2,1,2)模型有非常小的错误率，而 *ARMA*(1,2)会在一些时间点上发生明显错误预测。同样，在早上 9:05 左右高峰时段边界，*ARIMA*(2,1,2)还准确地预测了拥堵后的速度，而 *ARMA*(1,2)明显产生较大的误差。结果表明，在高峰时段，*ARIMA*(2,1,2)的预测精度比 *ARMA*(1,2)高, 假设二成立。

2.2　ANN 预测模型

2.2.1 BP 神经网络

BP 神经网络是一种按照误差逆向传播算法训练的多层前馈神经网络**错误!未找到引用源。**，它的基本思想是梯度下降法，利用梯度搜索技术，以期使网络的实际输出值和期望输出值的误差均方差为最小。*BP* 神经网络学习过程由正向传播和反向传播组成，在正向传播过程中，输入信息从输入层传经隐含层逐层处理，并转向输出层，每一层神经元的状态只影响下一层神经元的状态，处理后产生一个输出，并获得实际输出值和期望输出值的均方误差。如果在输出层不能得到期望输出值，则转入反向传播，将误差信号沿原来的连接通路返回，通过修改各神经元的权值，使得误差信号最小。

2.2.2 BP 神经网络建模步骤

a)确定网络结构：确定网络层，每层的神经元数量和激活函数。

BP 神经网络由输入层，隐含层和输出层组成，本文选取 $V_{实时}$，$V_{历史}$，*time*，*PH* 四个参量作为输入层，交通流速度作为输出层，为了确定隐含层中神经元的数量，实验选取了 3 到 15 个神经元进行实验对比，选取 4 个路段 11 月 1 日至 12 月 22 日的数据作为训练集，12 月 23 日至 12 月 29 日的数据作为测试集。从图 7 可以发现，具有 12 个隐含神经元的 *BP* 网络具有最佳性能，所以，将隐含层节点个数设置为 12。从而建立 4 个输入节点和 1 个输出节点，12 个隐藏神经元，隐含层激活函数为 *Sigmod*，输出层激活函数为 *Relu* 的神经网络来进行训练和测试。

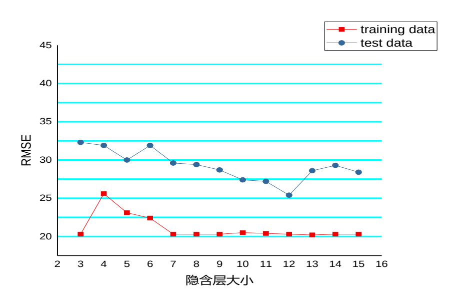


图 7　不同隐含层大小的神经网络性能

b) 初始化隐含层神经元阈值 θ_j，输出层神经元阈值 θ，在相同的采样数据下，设置最小误差 *E*，学习率 η，最大训练次数 *n*。

c) 根据输入样本计算隐含层输出和输出层输出。

d) 计算实际输出和期望输出的均方差，如果误差值满足要求，则结束网络训练。

$$E = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \mu)^2} \quad (13)$$

N 是样本数量，y_i 是预测值，μ 是期望值

e) 重复调整每个神经元的权重值，通过重复步骤 c)~e)使 *E* 达到误差范围要求。

3　实验结果及分析

3.1　评价指标

在该模拟实验中使用均方根误差作为评价指标，如式(14)所示。

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2} \quad (14)$$

其中：X_i 和 Y_i 分别代表实际速度和预测速度，*n* 代表预测的次数。

3.2　实验对比及分析

在三次实验中 *PH* 分别被设定为 4，5，6。当 *PH=4* 时，分别对高峰段和非高峰段的交通流速度进行预测分析，从图 8(b)中可以发现，在非高峰时段，*ARMA*(1,2)、*ARIMA*(2,1,2)和 *TS-ANN* 的预测走势总体一致，但相比于 *ARMA*(1,2)模型，*TS-ANN* 能更好反映交通流速度。在高峰时段，*ARIMA*(2,1,2)模型能够更地拟合出时序数据的线性关系，从而提高预测的准

确性。而本文的 *TS-ANN* 较 *ARIMA*(2,1,2)表现出显著的预测性能，但实际中还是有一部分时刻的交通流速度没能很好的预测到。未能预测到的主要的原因其一是该时刻点处于高峰时段发生和结束边界，其二是实验中使用的原始时序数据有限，本实验只使用了二个月的交通流速度信息，如果能采集到更多的时序数据，则能更好的拟合实际值。

当 *PH=6* 时，从图 9(*b*)发现，在非高峰时段，*ARMA*(1,2)、*ARIMA*(2,1,2)和 *TS-ANN* 的预测的数据走势总一致，且 *ARIMA*(2,1,2)比 *ARMA*(1,2)的预测性能较高，跟假设一的结果一致，但是并没有显著性的差距。而 *TS-ANN* 能更好的显示出道路的真实运行速度。在高峰时段时，*ARMA*(1,2)模型的拟合效果最差，且在部分时间点反映出较大的预测误差，而本文的 *TS-ANN* 的预测结果最准确，在高峰时段，*RMSE* 最低可达到 4.35%。

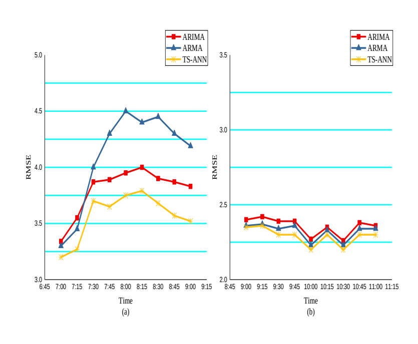


图 8　短时预测性能对比(*PH*=4)　(*a*)高峰期(*b*)非高峰期

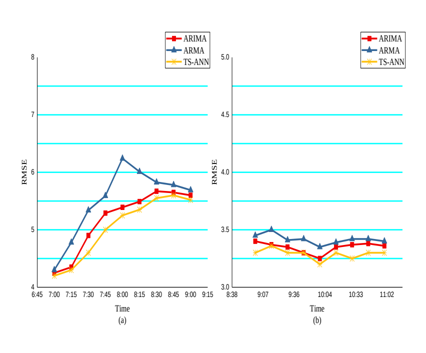


图 9　短时预测性能对比(*PH*=6):(*a*)高峰期(*b*)非高峰期

当 *PH=5* 时，如图 10 显示出不同时刻真实值与预测值的拟合情况，可以看出，在 6:00 到 6:50 期间，*ARMA*(1,2)模型和 *TS-ANN* 跟实际值的相似程度最高，而 *ARIMA*(2,1,2)模型在部分时刻出现较大的计算偏差，比如在 6:15，*ARIMA*(2,1,2)的预测值是 *27.6km/h*，而实际值为 *43.6km/h*。在 6:50 左右的早高峰时，*ARMA*(1,2)模型并没有捕捉到道路的真实状态，直到 7:13 左右时才反映出道路的拥堵状态，而 *TS-ANN* 用非常小的误差率预测拥堵的开始。在早高峰结束时，*TS-ANN* 也成功预测到道路的真实状态，结果表明，本文的 *TS-ANN* 在短时交通流速度预测中可以很好地处理预测范围以及预测时间的影响，模型具有一定的适应性。

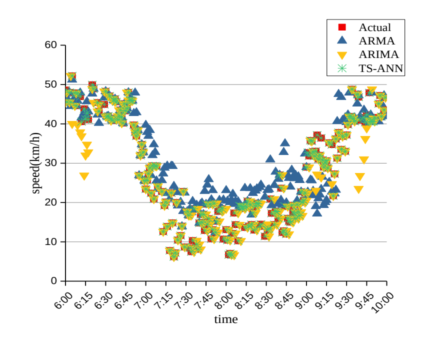


图 10　不同模型预测值与真实值比较

4　结束语

交通道路的变化过程具有复杂性和不确定性，单纯使用某一种模型难以得到满意的预测结果。本文提出一种城市道路交通流速度预测模型，该模型充分使用了时间序列模型和人工神经网络在线性和非线性建模的优势，在可预测性分析的基础上，分别在实时数据和历史数据中选取一组时序数据进行时间序列建模预测，对二者的预测值通过 *BP* 神经网络进行融合调整，从而得到最终的预测结果，在一定程度上改善了短时交通流速度的预测效果。但本文并未考虑到天气和突发事件等因素对路段速度的影响，因此对该类因素的分析将是本文下一步的主要研究工作。

参考文献：

[1] 冯凯. 城市智能交通系统的发展现状与趋势 [J]. 环球市场信息导报, 2017, 6 (1): 111-112. (Feng Kai. Development status and trend of urban intelligent transportation system [J]. Global Market Information Guide, 2017, 6 (1): 111-112.)

[2] 孙晓亮. 城市道路交通状态评价和预测方法及应用研究 [D]. 北京: 北京交通大学, 2013. (Sun Xiaoliang. Urban road traffic state evaluation prediction: a new scheme applications [D]. Beijing: Beijing Jiaotong University, 2013.)

[3] 马建, 孙守增, 芮海田, 等. 中国交通工程学术研究综述 [J]. 中国公路学报, 2016, 29 (6): 1-161. (Ma Jian, Sun Shouzeng, Rui Haitian, *et al.* Review on China’s traffic engineering research progress [J]. China Journal of Highway and Transport, 2016, 29 (6): 1-161.)

[4] Lai Qingying, Liu Jun, Luo Yongji, *et al.* A hybrid short-term forecasting model of passenger flow on high-speed rail considering the impact of train service frequency [J]. Mathematical Problems in Engineering, 2017, 2017 (3): 1-9.

[5] Moreno E G, Romana M G, Martínez O. A first step to diagnostic of urban transport operations by means of GPS receiver [J]. Procedia Computer Science, 2016, 83 (44): 305-312.

[6] Cao Hongmei, Han Feng. The urban arterial traffic flow forecasting based on BP neural network [C]// Proc of the 4th International Conference on Instrumentation and Measurement, Computer, Communication and Control. 2014: 393-397.

[7] Zhang G P. Time series forecasting using a hybrid ARIMA and neural network model [J]. Neurocomputing, 2003, 50 (1): 159-175.

[8] Li Kuilin, Zhai Chunjie J, Xu Jianmin. Short-term traffic flow prediction using a methodology based on ARIMA and RBF-ANN [C]// Proc of Chinese Automation Congress. 2017: 2804-2807.

[9] Box G E P, Jenkins G M, Reinsel G C, *et al*. Time series analysis: forecasting and control [M]. Hoboken: Wiley, 2015.

[10] Anokye R, Acheampong E, Owusu I, *et al*. Time series analysis of malaria in Kumasi: using ARIMA models to forecast future incidence [J]. Journal Cogent Social Sciences, 2018, 4 (9): 1-15.

[11] Bogl M, Aigner W, Filzmoser P, *et al*. Visual analytics for model selection in time series analysis [J]. IEEE Trans on Visualization and Computer Graphics, 2013, 19 (12): 2237-2246.

[12] Medel C A, Salgado S C. Does the BIC estimate andforecast better than the AIC [J]. Revista de Analisis Economico–Economic Analysis Review, 2013, 28 (1): 47-64.

[13] 王会青，郭芷榕，白莹莹. 一种基于 BP 和朴素贝叶斯的时间序列分类模型 [J]. 计算机应用研究，2019, 36 (8) . http://www. arocmag. com/article/02-2019-08-020. html (Wang Huiqing, Guo Zhirong, Bai Yingying. Time series classification model based on BP and Naive Bayes [J]. Application Research of Computers, 2019, 36 (8) . http://www. arocmag. com/article/02-2019-08-020. html)